OSBR.ca
The Open Source Business Resource

FEBRUARY
2008

FEBRUARY 2008

On February 8, as the Open Source Definition and the Open Source Initiative entered their second decade, Bruce Perens published his reflections in State of Open Source Message: A New Decade for Open Source (http://perens.com/works/articles/State8Feb2008/). He states his intent was to provide "another way of talking about Free Software, tailored to the ears of business people". Seeing that Gartner is predicting that "by 2011 at least 80% of commercial software will contain significant amounts of open source code", it would appear that the open source message has succeeded in reaching the ears of business people.

While awareness of open source code has reached a certain level of maturity, the innovation possibilities extend beyond the code itself. As you'll see in this issue of the OSBR, awareness of open "data" is still in its infancy, resulting in many untapped opportunities.

Tracey Lauriault and Hugh McGuire from Citizens for Open Access to Civic Information and Data describe current restrictions on Canadian public data and provide examples of potential benefits as access to this data becomes open. Joseph Potvin examines how Canadian copyright draws upon both the British and French traditions and the impact for those who deal with source code, data, and databases.

Jordan Hatcher from the Open Data Commons Project introduces the Public Domain Dedication & License and the CCZero tool which allow data and databases to be placed into the public domain. Ismael Peña-López from the Open University of Catalonia examines the benefits of a Personal Research Portal to knowledge workers.

I'd like to encourage readers to take advantage of the tools in the OSBR website. You can use these tools to post comments, notify colleagues, send emails to authors, access printable versions, access article metadata and search for additional works. All content is released under a Creative Commons license, meaning you are free to link, discuss, and reprint any content as long as you provide attribution. If you blog or review any articles, let the author know--they will appreciate the exposure.

We welcome suggestions for themes of future OSBR issues (e.g. open source for geospatial applications, open source for health) and names of potential authors of insightful articles. Please send your suggestions via email to the Editor. As always, we look forward to your feedback and suggestions for improving the OSBR.

Dru Lavigne

Editor-in-Chief

dru@osbr.ca

Dru Lavigne is a technical writer and IT consultant who has been active with open source communities since the mid-1990s. She writes regularly for O'Reilly and DNSStuff.com and is author of the books BSD Hacks and The Best of FreeBSD Basics.

"Data constitute a critical national resource, one whose value increases as the data become more readily and broadly available."

(http://www.nap.edu/catalog.php?record_id=4871)

There is a global movement to liberate government-"owned" data sets, such as census data, environmental data, and data generated by government-funded research projects. This open data movement aims to make these datasets available, at no cost, to citizens, citizen groups, non-governmental-organizations (NGOs) and businesses. The arguments are many: such data spurs economic activity, helps citizens make better decisions, and helps us understand better who we are and where we are going as a country. Further, these data were collected using tax dollars, yet the government holds a monopoly which makes data available only to those able to pay the high access fees, while some data is not made available at all.

The open data movement is lagging in Canada as demonstrated by exorbitant fees for such basics as the data set of postal codes correlated to electoral districts. This data could be used for any number of civic engagement projects, but it costs thousands of dollars due to Statistics Canada's policies of cost recovery.

Currently, access to government data is hampered by four main factors: i) the high cost of available data sets; ii) arbitrary decisions about availability of data sets to the public; iii) restrictive licenses; and iv) inaccessible data formats.

Formed in 2007, Citizens for Open Access to Civic Information and Data is a loose grouping of academics, activists, and citizens concerned with promoting data liberation in Canada (http://civicaccess.ca).

The grouping includes lawyers, copyright experts, librarians, archivists, cartographers, engineers, communications activists, open source programmers, and new media designers. The two main objectives of CivicAccess are:

• encourage all levels of governments (e.g. federal, provincial, municipal) and sectors (e.g. health, environmental, education) to make civic data and information available to citizens without restrictions, at no cost, in usable open formats

• encourage the development of citizen projects using civic data and information

The long-term vision is a country in which citizens, specialists, professionals, academics, community groups and even businesses can work together, developing innovative information access and visualization tools, better decision-making models, and more tools responsive to the needs of the citizens. Liberating data will spur grassroots research on important social, economic, political and technical areas, currently hampered by lack of access to and high cost of civic data. Further, we want to link the debate about data to questions of government transparency and accountability, which pivot on access to accurate, reliable, and timely data.

But first, we need access to that data.

What are Civic Data?

Civic data are a public good, and more specifically, are "numerical quantities or other factual attributes generated by scientists, derived during the research process through observations, experiments, calculations and analysis (http://nsf.gov/pubs/2005/nsb0540/)".

It is also "facts, ideas, or discrete pieces of information, especially when in the form originally collected and unanalyzed" (http://www.archivists.org/glossary/), and also, from the Report of the National Science Board, "numbers, images, video or audio streams, software and software versioning information, algorithms, equations, animations, or models/simulations". Distinctions are made between raw or level 0 data and derived, refined, synthesized or processed data. Raw data are normally unprocessed; examples include digital signals from a sensor or an instrument (e.g. unprocessed satellite image, thermometer), facts derived from a sample collected for an experiment (e.g. blood sample, ice core), and facts collected by human observation (e.g. mine tailings, census). Computations and data manipulations are related to research objectives and methodologies. Refined or processed data are raw data that have been manipulated, undergone computational modeling, been filtered through an algorithm, sorted into a table or rendered into a map. In these cases, access to the models is as important as access to the output results of those data.

Civic data are the data created and maintained by public organizations and paid for by the public purse as part of the on-going day-to-day activities of governing. Public data can include crime data at the neighbourhood scale, the number of traffic violations for certain streets, election results, census data, road networks, non-private health data, government expenditure data, school board catchment area boundaries, aggregated test results, environmentally sensitive or contaminated areas, or basic framework map data that include census areas, administrative boundaries, postal code areas and geo-referenced satellite images. Framework data are particularly important as these are the foundational data sets upon which other datasets can be organized.

Civic data also includes those created as part of government funded research organizations such as the Social Sciences and Humanities Council of Canada (SSHRC) and the Natural Sciences and Engineering Research Council (NSERC) or any other outsourced publicly funded data and information creation activity.

Types of Open Data?

Some aspects of the open data movement (see also the Hatcher article in this issue) include:

• Open Access (OA, http://wikipedia.org/wiki/Open_access), which aims to end restrictive licenses on university research and data as seen in initiatives such as Open Access News (http://www.earlham.edu/~peters/fos/fosblog.html)

• data visualization projects which combine design and data in creative ways to make information more accessible, such as Gapminder (http://gapminder.org/)

• grassroots citizen projects using government data sets to improve cities and towns, such as FixMyStreet (http://www.fixmystreet.com/)

Civic Data Access in Canada

Access to civic data in Canada depends on how much money you have, to which organizations you are a member, and for what purpose you want to use the data.

If you are a university professor or tuition paying university student in Canada, access to data is quite good. This is largely the result of work done by the Data Liberation Initiative (DLI) which is a data purchasing consortium (http://statcan.ca/english/Dli/dli.htm).

DLI consortium members pay an annual subscription fee that allows their faculty and students unlimited access to numerous Statistics Canada public use microdata files, databases and geographic files. If you are a student or teacher in Ontario, you may access data from the new Ontario Data Documentation, Extraction Service and Infrastructure Initiative (ODESI, http://odesi.uoguelph.ca/wiki/index.php/Main_Page) which will target Statistics Canada datasets, datafiles from Gallup Canada and other polling companies, public-domain files such as the Canadian National Election Surveys, and selected files from the Inter-University Consortium for Political and Social Research (ICPSR).

Both the DLI and ODESI provide access to a small subset of Canadian citizenry. Their license is very specific about who authorized users are, exclusivity, and how data products cannot be used such as "in the pursuit of any contractual or income-generating venture either privately, or under the auspices of the educational institution" (http://www.statcan.ca/english/Dli/caselaw/assess.htm).

If you work for a government, access to data varies depending on which department and level of government you are in, the rationalization you have for acquiring that data, and the budget your department or section has. For instance, Environment Canada shares its data quite openly, as does Natural Resources Canada via the GeoConnections program (http://www.geoconnections.org/). While much data comes with a price tag, many data sets are free such as those under the GeoGratis.ca (http://geogratis.cgdi.gc.ca) or the Geobase.ca (http://geobase.ca/) programs. In fact, Geobase.ca has one of the most progressive data licensing programs (http://www.geobase.ca/geobase/en/licence.jsp) so far seen in the Government of Canada.

At the Canadian provincial or city scale, things start to get confusing as licenses differ, as do cost recovery and access policies. Land Information Ontario (LIO) has many data sets in their downloadable catalogue (http://tinyurl.com/yufhn5); however, this data is only available through a Government of Ontario Intranet or between and among members of the Ontario Geospatial Data Exchange (OGDE, http://www.lio.gov.on.ca/en/Exchange.htm). Municipalities suffer from very restrictive or non-existent data sharing policies that are not uniform across departments.

As an example, the City of Ottawa has different categories of clients for its GIS data:

• category A, internal municipal clients: no charge for data and rarely require a license agreement

• category B, external municipal clients: are charged a fee to reflect the staff resources consumed in the preparation of the data and sometimes require a license agreement

• category C, external groups needing data for specific projects: are usually charged the same fee as category B clients and must also enter into a signed data license agreement naming a specific project or use

• category D, external groups wishing to commercially market the data: category D clients are expected to pay a fair market rate for any data they want to commercialize

"for all requests it is expected that the client can demonstrate a legitimate use of the data. This provision ensures that staff resources are not unduly expended on frivolous requests."

"Additionally, the license must refer to a specific project or use as this helps the City track how the data is being used and by whom."

There is no "citizen" category. How you can use, re-use, and represent data are quite restrictive. It would seem logical to have data discoverable and accessible via a data portal. This would result in the City not having to work so hard to micro manage the use of our public data.

Things get really confusing when different levels and departments of government repeatedly sell each other the same data sets with public money. Governments do not have intra-governmental data portals that centralize data acquisitions and share data assets amongst public servants. Duplication of effort and multiple layers of bureaucracy and accounting could be done away with by simply making all the data free to not only citizens but also their governments!

If you are from an NGO, data access is cost prohibitive. Many small NGOs pool their resources and develop data purchasing consortia such as the Canadian Council on Social Development Community Social Data Strategy (http://www.ccsd.ca/subsites/socialdata/home.html).
However, like the DLI, these entities remain closed and exclusive shops.

Statistics Canada allows a variety of companies to resell civic data (http://www.statcan.ca/english/reference/data.htm) and has also licensed a number of civic data value added distributors (http://www.statcan.ca/english/reference/value.htm).

As a citizen, you have access to incomplete data sets from the Depository Service Program (http://dsp-psd.pwgsc.gc.ca/Reference/guides-e.html) available to you in public libraries.

These are suitable for high school projects but not for public participation in a democracy. What we really need is a concerted lobby in Canada that will free public data.

Why Free Civic Data?

In a wider, less technical sense, "data" are what we use to make decisions, so they are a public good. We use data sets to make decisions about how we as individuals should act, and how we as a society ought to do things. All the rules that govern our societies, from agricultural practices to cooking, to our law systems and social interactions, are the result of our interpretation of the interaction between different data sets over time.

Our ability to collect, analyze and interpret these data, and to make decisions based on them, is what gives humans our particular ability to solve societal problems such as food shortages, disease infestations, and resource depletion.

Democracy has a number of fundamental ideals, including free speech, free press, transparency of government, separation of powers, rule of law, public education, and free markets. All these principles are based on openness of information, or openness of data. In a sense, the basis of democracy is to open up the decision-making process to everyone.

By opening data to more people, you get more interpretations, more proposals of different solutions, better decisions about the best solutions, and in the long run, more successfully-solved problems.

We have reached a time when the cost to share datasets is no longer cost prohibitive.

The processing power available on a desktop computer can do an enormous amount with even large datasets. Skilled designers have the ability to interpret, re-design, repackage, and display data in new and important ways, and the social web allows others to contribute to that process.

Transparency and accountability are essential elements of a functional participative democracy, and access to data and information is imperative. Transparency increases as quality data are widely and freely disseminated. Government and the private sector often miss important types of analyses, particularly local, cross boundary or jurisdictional research.

For instance, it is cost-prohibitive and technically difficult for a community group to discover and access neighbourhood-scale data from different levels of government to conduct any kind of local community market or demographic analysis. An entrepreneur developing a business plan for a company to operate in four cities in two provinces would quickly discover restricted access to the basic data and information required to understand their market niche, clients, and competitors.

The basic digital data and information upon which we depend are rarely accessible, rarely interoperable, rarely in open formats, and are often prohibitively expensive. Moreover, regressive licensing regimes impede the sharing of data, or worse, there are no licensing regimes at all, which leaves citizens at the whim of the decisions of public servants. This is particularly true at the municipal and school board levels where a lack of clear guidelines often means no access to data for fear of releasing the wrong thing. For Canadian citizens this means that much innovation and knowledge is being thwarted.

Worse, we often are forced to pay exorbitant prices for data to study important issues such as poverty (http://datalibre.ca/2007/11/30/paying-for-data-to-study-poverty/), homelessness (http://datalibre.ca/2008/01/18/housing-and-homelessness-data-in-canadian-cities/) or to assess the cost to the health care system of poor air quality (http://www.oma.org/phealth/icap.htm).

Civic Data Projects

Wikipedia was launched in 2001, and in seven years has displaced Britannica, the gold-standard English language reference encyclopedia since 1768. Wikipedia has more articles, is more up-to-date, and, while the accuracy of the information in Wikipedia is a constant work-in-progress, Nature's December 2005 study of scientific articles in the two encyclopedia found the accuracies to be roughly equivalent (http://tinyurl.com/yotjyh). Wikipedia is the most useful encyclopedia in the world, if, by useful we mean, "the encyclopedia that most people use."

We are beginning to see more examples of civic projects. One example gets right to the nitty-gritty of municipal politics: potholes. Launched in February 2007, the UK project FixMyStreet.com "is a site to help people report, view, or discuss local problems they've found to their local council by simply locating them on a map." The project targets such problems as potholes, broken streetlights, and graffiti. It has revolutionized municipal maintenance planning by putting the data collection into the hands of citizens and opening up the planning and decision-making process to many concerned citizens. Problem reports are there for all to see, providing municipal councils more incentive to fix the problems.

Another amateur project that turns a light on the political process itself is howdtheyvote.ca, which tracks how Canadian members of parliament vote on individual bills -- information that should be fundamental to our understanding of our representatives in Parliament.

Crimereports.com is a US site built to help citizens get more information about the locations and frequencies of crime incidents in their cities.

These examples of progressive initiatives suggest that we are in the early days of the movement towards opening up government data. Open data allows citizens to build tools that can address issues important to them. More tools of civic engagement through data are starting to appear on the web, and there is much to be done.

What is CivicAccess.ca Doing?

Civicaccess.ca is about liberating public data from public institutions and finding new ways to make data accessible and useful. Individual members are doing incredible things. However, as a collective we have not tackled any big projects. We provide a mailing list (http://lists.pwd.ca/mailman/listinfo/civicaccess-discuss) with over 150 members across the country that exchange information on issues, innovations, projects and ideas.

The authors of this paper also co-author DataLibre.ca, a CivicAccess.ca inspired blog, to fill a void on this topic. Its readership has been increasing and we are seeing traffic coming from key players in the open access movement, the open data and open source communities, along with members from library and archives associations.

Ultimately, CivicAccess.ca is firing up the conversation on access to public data in Canada and we hope to discover and support the creation of innovative open public data projects. So come and join us!

Conclusion

Innovation comes from many drivers and sources, but there are two essential prerequisites: a problem in need of solving, and information and data. With a few other ingredients such as intelligence, creativity, and resources, innovation will occur. But the fundamental ingredients in innovation are always human desires to improve something, and figuring out, based on information, how to improve it.

Solving problems is one fundamental role of governments. By opening up civic data, and allowing citizens and citizen-groups to participate in problem solving, we believe that we will start to see more innovative and better solutions to the problems facing society.

Doing any form of research requiring cross jurisdictional civic data sources that cross domains, sectors and topics is very difficult in Canada. We have discussed the underlying reasons, examined some of the many bottlenecks and roadblocks, and highlighted examples of some progressive initiatives.

The technological solutions to provide free access to Canada's civic data are readily available and relatively inexpensive (http://library2.usask.ca/gic/v2n4/mcmahon/mcmahon.html). What is more difficult is finding the political will to make our civic data public.

Recommended Resources

Klinkenberg, Brian, The True Cost of Spatial Data in Canada
http://tinyurl.com/29f6de

McMahon, Ronald C., Saskatchewan Bureau of Statistics, Cost Recovery and Statistics Canada
http://library2.usask.ca/gic/v2n4/mcmahon/mcmahon.html

UK Guardian Free Our Data Campaign
http://www.freeourdata.org.uk

PodCast about CrimeReports.com
http://itc.conversationsnetwork.org/shows/detail3459.html

Stephenson, W. David, Let my data go! The Case for Transparent Government
http://stephensonstrategies.com/speeches/let-my-data-go-the-case-for-transparent-government/

Tracey P. Lauriault is a geographer. She is a member of the Senior Advisory Committee for an On-line Health Data and Community Mapping Portal, the Geographic and Numerical Information System (GANIS), and a research associate with Acacia Consulting and Research. Her research includes access and preservation to scientific data, olfactory cartography, transdisciplinary research, community mapping, homelessness, the application of geomatics technologies, cybercartography and infrastructures. She co-edits datalibre.ca, a blog about public access to government data in Canada.

Hugh McGuire is a Montreal-based writer, web developer and free data activist. He is the founder of LibriVox.org, a volunteer Internet project with the objective of making free audio versions of all books in the public domain, now the most prolific audio book maker in the world. He co-edits datalibre.ca, a blog about public access to government data in Canada.

"It is, I think, an elementary principle of copyright law that an author has no copyright in ideas but only in his expression of them. The law of copyright does not give him any monopoly in the use of the ideas with which he deals or any property in them, even if they are original. His copyright is confined to the literary work in which he has expressed them. The ideas are public property, the literary work is his own."

Justice Thorson P.
in Moreau v. St. Vincent

One of the essentials of a healthy and democratic economy is that the rules of engagement should be understandable to people whose interests are affected by them. There are two aspects of copyright law to which all software and database professionals ought to be familiar, but usually are not. First, they should know that factual data listed in an obvious structure is not covered by copyright law, no matter how much work went into collecting it. Second, they should know that under the Canadian Copyright Act, programming code is considered to be a "literary work".

This article outlines how copyright is related to source data and source code, and why this is important to both technical and business professionals in the field.

Origins and Implications of Copyright

After Johann Gutenberg invented the printing press in 1440, it became easier for people to disseminate heretical and seditious works, challenging both church and state. In order to control what was being said, Henry VIII of England in 1538 invoked royal prerogative, on questionable constitutional grounds, to establish printing patents, as a form of censorship.

Then, by a royal charter in 1557, the Stationers' Company was created by the British Crown to oversee a guild system in which the right to print a book was limited to members of the guild, who were the printers and sellers of books, not the authors (http://tinyurl.com/yu4oq6). The Stationers' Company charter declared:

"Know ye that we, considering and manifestly perceiving that certain seditious and heretical books rhymes and treatises are daily published and printed by divers scandalous malicious schismatical and heretical persons, not only moving our subjects and lieges to sedition and disobedience against us, our crown and dignity, but also to renew and move very great and detestable heresies against the faith and sound catholic doctrine of Holy Mother Church, and wishing to provide a suitable remedy in this behalf..." (http://www.hup.harvard.edu/catalog/ROSAUT.html).

Much has changed in 350 years, but copyright in Canada today should still be interpreted in its historical context, with attention to our own current Copyright Act (http://lois.justice.gc.ca/en/C-42/text.html), case law as it has developed through court decisions, as well as international conventions. Under Section 91(23) of the Constitution Act, 1867, the federal government was granted exclusive power to enact laws within Canada related to copyright. But Canada remained under British Copyright until 1921 when the Canadian Parliament passed its own Copyright Act. So, when the United Kingdom ratified the Berne Convention for the Protection of Literary and Artistic Works in 1887, they also ratified it on behalf of Canada. Canada only ratified the Berne Convention as a separate country in 1928. Even Canada's Copyright Act, enacted in 1921, and which came into force in 1924, was closely modelled on the English Copyright Act of 1911.

It is a common experience in bilingual and multilingual settings to encounter problems of confusion when semantic meaning gets lost in translation. In this case, we find something gained in translation with the French phrase droits d'auteur (http://en.wikipedia.org/wiki/French_copyright_law), which is evidently not "droit de copier". The English word copyright refers to a straightforward economic right to make copies of a work.

This is extended in the French droits d'auteur in a way that draws upon the European continental civil law (civiliste) tradition. It holds that the right of reproduction goes beyond the simple right to make new copies of a work, to the more complex notion of protecting the integrity and paternity of the work, because it is linked to the author's reputation in society.

Droits d'auteur does not refer strictly to the dollars-and-cents linkage between an author and the creative work. Instead, the work is considered to represent something about the author, whose dignity deserves protection, although the courts have been clear that the author cannot be the judge in their own cause. The Théberge v. Galerie d'Art du Petit Champlain Inc. case provides an excellent description of these concepts (http://www.canlii.org/en/ca/scc/doc/2002/2002scc34/2002scc34.html).

Copyright in today's socio-economic milieu has come to take on a different purpose. In today's global digital age of peer-to-peer computing, data warehouses, mash-ups, wikis and free/libre licensing, one must also realize that the key concepts and definitions in each country remain a little different, which inevitably leaves much room for confusion.

Even within our own Canadian legal context, it is challenging for database and software professionals to steer clear of misunderstanding due to the conceptual differences between the droit d'auteur continental civiliste tradition that emphasizes personal reputation, and the English copyright tradition that emphasizes artistic and literary works as articles of commerce.

In Canada, Parliament has sought to draw upon both traditions, providing a legal scenario that some would say holds true to the comment that Canadian publisher Stuart Keate made about the country in general: "in any world menu, Canada must be considered the vichyssoise of nations, it's cold, half-French, and difficult to stir." But the duality Canadian legislators have tried to accommodate is: i) a right that is centred on the reputation of the person of the author (denoted "moral right"); and, ii) a right centered on the economic role of the work as an object of commerce. Thus, in addition to having the right to assert controls and issue licenses over publishing, production, reproduction or performance of a original work in material form, in whole or in part, Canada's Act provides that the author has a right to the integrity of the work and the right, where reasonable in the circumstances, to be associated with the work as its author by name or under a pseudonym, and the right to remain anonymous.

Section 28.2(1) specifies, however, that "the author's right to the integrity of a work is infringed only if the work is, to the prejudice of the honour or reputation of the author" as a result of the work being "distorted, mutilated or otherwise modified" or "used in association with a product, service, cause or institution".

In Canada, the reputation (moral) rights of an author can be waived, but not transferred through assignment or sale, whereas copyright can be sold or assigned to a person or entity other than the original author.

Implications of Copyright for Data and Database Professionals

The boundary line regarding the applicability of copyright law to data was clarified in a 1997 case at the Canadian Federal Court of Appeal (TeleDirect Inc. v. American Business Information Inc., http://tinyurl.com/2jle3x). In his decision, Judge J.A. Denault explained: "Under subsection 5(1) of the (Copyright) Act, copyright subsists not in a compilation of data per se, but in an original work... the selection or arrangement of data only results in a protected compilation if the end result qualifies as an original intellectual creation". He reiterated a United States Supreme Court decision (Feist Publications, Inc. v. Rural Telephone Service Co. Inc., http://tinyurl.com/yg34pv) which found that listings of routine factual data, such as names, towns and telephone numbers in a telephone directory, are uncopyrightable facts, because they are not selected, coordinated, or arranged in an original way.

The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS, http://en.wikipedia.org/wiki/TRIPS) also states: "Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself."

Further, Subsection 2.1(2) of the Copyright Act states that the "mere fact that a work is included in a compilation does not increase, decrease or otherwise affect the protection conferred by this Act in respect of the copyright in the work". The courts have consistently found that the amount of effort required to collect and manage the information is not a criterion for copyrightability.

What does this mean for database professionals? It provides that while published financial reports are copyrightable, the generic compilation of source data behind them is not. For example, the Public Accounts of Canada report is distinguished from the Accounts of Canada maintained in the Central Financial Management and Reporting System, in the Financial Administration Act. The Accounts of Canada is a database containing records of expenditures, revenues and other payments to and from the Consolidated Revenue Fund, as well as records of assets, liabilities and reserves. In contrast, the Public Accounts is "a report...prepared by the Receiver General for each fiscal year and...laid before the House of Commons...in such form as the President of the Treasury Board and the Minister may direct". The report not only includes summary statements of financial expenditures and revenues, assets liabilities and contingencies; it also contains the opinion of the Auditor General, and additional information to communicate the financial position of the government (http://laws.justice.gc.ca/en/F-11/index.html).

Rights in the other constituent parts of a database also warrant consideration. Copyright title to generic documentation of the source data model and metadata schema may be held by an international standards body, such as the International Public Sector Accounting Standards Board (http://www.ifac.org/Public Sector/).

Similarly, title to other discrete parts of the database would be determined according to what organizations the database analysts/architects worked for, their terms of employment with those organizations, and the terms of the federal contracts under which the work was performed.

From the above, it is therefore important for database professionals to consider copyright title to technical documentation of the source data model such as entity-relationship diagrams and UML (unified modeling language) diagrams, to the implemented database tables, indices and functions, and even to the data entry forms, queries, and output views. The above also suggests a way to filter out some potential copyright complications when obtaining data from other third party compilations, in order to ensure a simpler, more practical business rules scenario for one's own users. The key is to populate one's own database with flat comma separated value imports of selected pure source data.

Implications of Copyright for Source Code Professionals

Under the Copyright Act, a computer program is a literary work/oeuvre litteraire, of a type that occurs as "a set of instructions or statements, expressed, fixed, embodied or stored in any manner, that is to be used directly or indirectly in a computer in order to bring about a specific result". Similarly, the international TRIPS agreement states that "computer programs, whether in source or object code, shall be protected as literary works under the Berne Convention (1971)". Accordingly, an expression in C such as:

```
#include <iostream>
int main()
{
std::cout << "Hello, world!\n";
}
```

or in Ruby:

```
for i in 1..1
puts "Hello World!"
end
```

has the same essential characteristics, in terms of information management and intellectual rights, as the English expression in pre-formatted text such as:

Print: "Hello World!"

It is important for any organization to maintain a consistent approach to written works, whether the intended readers are people or machines. Just as the text you are reading presently is also machine-readable via optical character recognition technology, the source code for the browser you are using is readable by someone fluent in the given programming language. Under the Copyright Act, they are identical.

In this light then, it does seem out-of-place to treat the acquisition of computer programming code like the acquisition of hard drives, as if one should consider stories to be similar to books. The methods of commerce applied to trade in licences for units of restricted-access software since the early 1990s have made it seem common sense for procurement professionals to treat software programs in terms of commodity units. But when one thinks of programming code composed under contract, or in-house by one's own employees, or code that is downloaded from external sources under free/libre/open license terms, it quickly becomes apparent that each of these are acquisitions of streams or compilations of information, like reports. Spending for programming code that is prepared under contract is accounted for under professional services; and when the code is written in-house, the money shows up as salaries.

On the contrary, payments to vendors for unit licenses are really rental and support fees in exchange for the installation or use of that software, which is usually bundled with financial services fees and other services fees. The latter are not acquisitions of software copyright. Accounting and acquisitions policies and practices that perceive this genre of literary work as commodity units rather than as streams or compilations of information, can be misleading.

It is psychologically tempting for end-users to think they actually own a copy of software, especially when it is delivered to them on physical media. But all they have is permission from the owner(s), through a license, to have a copy of the information, and to read it or have a machine read it. People talk in a colloquial manner about owning a copy of some software that they paid a license for.

The gap between owning a license and owning a commodity becomes quickly apparent as soon as they want to give it to somebody else. Under RENT (Restrictive/Exclusive/Negotiated Title) licenses, they cannot. Under FLOW (Free/Libre/Open Work) licenses, they can, within certain conditions. None of these licenses transfer any ownership of physical assets or intellectual rights to the copy holder.

In his Principles of Economics (http://www.econlib.org/library/Marshall/marP.html), Alfred Marshall observed that: "The distinction between public and private property in knowledge and organization is of greater importance than that between public and private property in material things; and partly for that reason it seems best sometimes to reckon Organization apart as a distinct agent of production."

A hundred years earlier, Thomas Jefferson wrote a letter to Isaac McPherson to articulate the practical distinction between public/private property considerations in relation to intellectual versus material things:

"If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me."

Jefferson's emphasis that the possession of intellectual things can be infinitely concurrent, while the possession of material things is ultimately exclusive, even when held in common, is obviously critical to any consideration of licensing and contracting.

The confusion over whether a software application is a commodity, like a book, or information, like a story, even shows up in formal accounting rules. The Canadian Institute of Chartered Accountants and its counterpart professional bodies internationally have, for twenty years, permitted the capitalization of software spending. This means the amount spent can be depreciated in the same way that material goods can be, to reflect declining market value due to wear and tear.

Professor Charles Mulford and Jack Roberts at the Georgia Institute of Technology recently analyzed how the capitalization of software expenditure causes financial reports to significantly overstate earnings for the fiscal year in which the money is spent, and then, through amortization, to cause earnings to be understated in subsequent years (http://mgt.gatech.edu/downloads/2006/ga_tech_software_dev_2006.pdf).

Finding that the majority of software development companies do not capitalize software spending, and that amongst firms where it is done, the methods are arbitrary, they recommend that accounting standards bodies should revoke the provisions that permit this practice. They propose that software development costs should be returned to the pre-1980s treatment as research and development, which is expensed. Such a step "would be more closely aligned with the realities of the software industry today".

The Main Points

Copyright began as a means of censorship, but eventually became grounded in a general rights framework, which, in Canada, draws upon both British and French legal concepts. If you work with databases or software, it is important that you understand how copyright law affects your rights today, so that you can make informed licensing and contracting choices.

Database professionals should be aware that basic facts displayed in an obvious structure are not covered by copyright. However, technical documentation, implemented database tables, indices and functions, data entry forms, queries, and output views are typically covered by copyright. The application of copyright should be considered by parties to contracts involving the use of or creation of databases.

Software professionals also need to pay attention to what software is. Confusion is widespread regarding whether software is a commodity or information. Misunderstanding on this point leads to a variety of errors in business judgement, that extend all the way to errors in financial reporting.

An initial overview of this paper was delivered by the author as the keynote presentation at the 2007 Annual General Meeting of the Statistical Society of Ottawa.

Joseph Potvin is an economist who has worked in public, private, academic and community organizations in several countries, with degrees from McGill (Canada) and Cambridge (UK).

16

"It won't be long before open access is old hat, taken for granted by a new generation of tools and services that depend on unrestricted access to research literature and data. As those tools and services come along, they will be the hot story. But historians will note that they all depend on open access and that open access was not easily won."

Peter Suber
(http://poynder.blogspot.com/2007/10/basement-interviews-peter-suber.html)

Free and libre/open source software (F/LOSS) movements have spawned similar solutions in many other contexts, each at differing stages of development. As F/LOSS enters the routine and familiarity of middle age, the open content movement--open source for non-software copyright and best embodied by the work of Creative Commons (http://creativecommons.org)--has just graduated university and is getting a feel for the world. Even younger is the open data movement, whose legal tools have just started to come online.

Open Data

Some may be surprised to learn that data and databases are not a "rights free" area where no intellectual property rights (IPRs) apply. For example, the Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPs) requires that members of the World Trade Organisation, including Canada, the United States, and the UK, provide legal protection for databases (http://tinyurl.com/2ecm27).

Rights covering databases can include:

• copyright: both for the selection and arrangement of the database contents and over the contents of the database itself (the data), though factual information will generally not be protected by copyright

• database rights: the European Union's Database Directive (http://tinyurl.com/2ywnna and http://www.nfais.org/publications/white_papers_2.htm) requires member states to implement a "sui generis (http://en.wikipedia.org/wiki/Sui_generis) database right" covering the extraction and re-utilisation of the contents of protected databases (**Editors note:** there is no North American equivalent to this directive)

• contract: contractual obligations about what users can and can't do with a database and its contents can also be used to provide for protection

• other rights: rights such as trade secrets and laws of unfair competition can also protect databases

This rights thicket protecting databases and data can form a significant obstacle for the use and re-use of data. This is true for both the scientific community wishing to expand knowledge through use of others' data and for the Internet and research community with aims to enable the semantic web.

Open Data Commons

With the funding and support of the information management company Talis, the Open Data Commons project (ODC) was founded in the autumn of 2007 to provide legal tools for sharing data (http://www.opendatacommons.org/).
This project started through funding licence development by Jordan Hatcher (http://opencontentlawyer.com) and Dr. Charlotte Waelde of the University of Edinburgh (http://www.law.ed.ac.uk/staff/charlottewaelde_77.aspx). This resulted in the creation of the Public Domain Dedication & Licence (PDDL) legal tool which will be maintained by the Open Knowledge Foundation (http://okfn.org), a not-for-profit organisation promoting open knowledge.

The PDDL dedicates the data and databases to the public domain, a position that offers a wide degree of flexibility for users of data and helps freely enable semantic web projects based on using large amounts of data.

Open Data Projects

Many projects of interest to the F/LOSS sector involve open data. These include:

Neurocommons: a Science Commons project which integrates data in the neurosciences (http://sciencecommons.org/projects/data/nc_technical_overview).

CKAN: the Comprehensive Knowledge Archive Network (http://www.ckan.net/), a registry of open knowledge projects maintained by the Open Knowledge Foundation and analogous to the freshmeat site for F/LOSS software.

Open Street Map: this site collaboratively produces open geodata (http://www.openstreetmap.org/).

Freebase: "an open database of the world's information" containing data from Wikipedia as well as US Government information (http://www.freebase.com).

Open data is also of enormous importance in the scientific community, where access to research data brings up many of the same issues as open access to scientific publications. For an overview of open access, see Peter Suber's introduction at http://www.earlham.edu/~peters/fos/overview.htm.

Science Commons Protocol

Science Commons was founded in 2005 and works on a variety of projects investigating rights issues related to scientific research (http://sciencecommons.org).

These include access to published research papers, material transfer agreements, and Neurocommons, a project creating an open source knowledge management platform for biological research. Science Commons is a project of Creative Commons and is overseen by its board.

In December 2007, Science Commons released their Protocol for Implementing Open Access Data (http://sciencecommons.org/projects/publishing/open-access-data-protocol/). This protocol, written in the same style as a Request For Comment (RFC, http://en.wikipedia.org/wiki/Request_for_Comments), outlines a legal standard for open access to data based on three principles:

• the protocol must promote legal predictability and certainty

• the protocol must be easy to use and understand

• the protocol must impose the lowest possible transaction costs on users

Guided by these three principles and Science Commons' experience in maintaining their database FAQ on Creative Commons licences and data, they arrived at an approach that calls for waiver of relevant IPRs so that data could be treated as close to being in the public domain (without IPRs) as possible. Thus the protocol calls for waiver of:

• copyright

• the sui generis database right in the European Union mentioned above and similar protections

• implied contract rights and rights in tort (http://en.wikipedia.org/wiki/Tort) or delict (http://en.wikipedia.org/wiki/Delict) such as unfair competition or trade secrets

This protocol gets enforced through the use of an "Open Access Data Mark", which will be managed by Science Commons and the sister organisation Creative Commons. They will limit use of the mark to licensing schemes that comply with the protocol, so that users can be assured that the data labeled with the mark meets the criteria of waiving IPRs. The Science Commons protocol thus sets a standard that any licensing scheme can implement.

Implementation in Open Data Commons

In implementing the Science Commons protocol, the ODC project set goals of:

• making the protocol international

• writing the legal document in plain language

• clearly stating what rights were and were not covered

From experience in the F/LOSS and open content communities, the ODC team thought it important to create a legal text as accessible as possible to its users. In terms of drafting style, ODC uses the same approach as the GPL (http://www.gnu.org/copyleft/gpl.html) in including such elements as a preamble, as well as the plain language approach of the Scottish implementation of the Creative Commons licences. Drafting efforts also drew heavily from the original Science Commons FAQ on databases, the Creative Commons unported licences, and the first generation Talis Community Licence. The result is the Open Data Commons Public Domain Dedication and Licence (ODC-PDDL, http://www.opendata commons.org/odc-public-domain-dedication-and-licence/) and an accompanying Community Norms document (http://www.opendatacommons.org/odc-community-norms/).

Waiving Rights

Copyright law as it relates to waiving copyright is unclear. No international treaties, such as the Berne Convention (http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html), set a standard for waiver of copyright. Indeed, it is unclear whether or not copyright can be waived in the United Kingdom, the physical home of the ODC project, and the same could be said to be true of the EU's sui generis database right. As a result, ODC decided on a two prong approach to implementation:

• waiver of database rights and copyright for jurisdictions that allow for it (see PDDL Sections 3.1 and 3.2)

• licensing of the rights for jurisdictions that do not allow for waiver (see PDDL Sections 3.3)

This approach accommodates the many different jurisdictional approaches to copyright law throughout the world while still setting the goal of waiving rights.

Moral Rights

Moral rights arise in some jurisdictions in connection with the creation of a copyrighted work. Because databases may attract copyright, they may also attract moral rights. These protect the rights of personal (as opposed to corporate) authors over their association with a work, including the right to be identified as the author of the work, and the right to object to derogatory treatment of the work.

Moral rights can be waived in the United Kingdom according to Section 87 of the Copyright Designs Patents Act. However, waiver of these rights may be impossible in many jurisdictions, especially those following the author's rights approach common in civil law jurisdictions.

Thus, while the PDDL waives moral rights in the work, Section 3.4 advises users that these rights may still nevertheless be present in the work.

Rights Addressed in the PDDL

The protocol clearly calls for the waiver or licensing of copyright and database rights, but these rights do not cover all the legal rights that could be potentially at issue in a database. The PDDL approach specifically excluded patent rights and trademarks. Patents could have been included in the same style as the GPLv3 which requires software patent holders to license these rights as they relate to GPLv3 licensed software. However, waiver or licensing of patent rights are not required under the protocol and would have greatly limited the utility of the PDDL. The exclusion of patent rights is included in Section 4.0 of the PDDL.

With regards to trademarks, Section 4.0 of the PDDL provides that the creator of the database should be able to maintain any marks associated with their own use of the database, even if they allow others to use the underlying database or data. In all cases, it was important that the provider of the data under the PDDL be placed in the same position as anyone else using the data.

Unfair Competition

The Science Commons protocol calls for waiver of unfair competition in Section 4.1. Unfair competition in US law, home of Science Commons, broadly refers to a group of distinctly different rights of action, including:

• trade secrets

• publicity rights

• trade mark claims

• passing off, a rights similar to trade-mark and based on the goodwill of a business

• deceptive advertising

• other kinds of unfair methods of competition

As you can imagine, the areas outlined above have a variety of different legal requirements and fit differing social policies. These rights of action, however, all involve using some aspect of a business without permission. Because the PDDL grants permission to use the data, specifically addressing this area did not seem to be required, and the ODC team hasn't heard any feedback that the PDDL does not adequately address unfair competition.

As an example, the law protects secret or confidential information, and trade secrets come under the umbrella term for unfair competition. If you use the PDDL and make your data available via the Internet, this database is no longer a secret and thus addressing this again would be redundant.

Database versus its Contents

The PDDL can be applied to both a database and its contents or data, or to only a database without covering its contents, as follows:

• database and contents: the entire database and data are free to use and re-use under the waiver and licence of the PDDL

• database only: any rights, such as sui generis database rights, that would accrue by creating, maintaining, and designing the database are waived and licensed under the PDDL, but the contents remain under other licences

The option to include the database and data or just the database elements is implemented by the definition of "Work" in Section 1.0 of the PDDL.

The option to cover only the database and not the database plus contents is present in the PDDL so that users creating databases with information under varied rights status, such as Freebase's use of both US Government data in the public domain and Wikipedia content under the GFDL, can apply the PDDL to only any rights present as a result of their creation of a database.

Community Norms Document

The PDDL works in conjunction with a non-binding Community Norms document. This document outlines in plain language a group of norms that users of a PDDL-licensed database should follow in order to create social obligations for data users. These norms include:

- reciprocal use of the PDDL: like the reciprocal or copyleft portions of the GPL or the Share Alike element in the Creative Commons licence, this norm asks users to release any changes to the database also under the PDDL

- attribution: this norms asks users to "give credit where credit is due"

- open formats: suggests the use of data formats that are accessible to all

- technical protection measures: technical restrictions such as digital rights management (DRM) are discouraged

The Community Norms document is flexible and adaptable to the norms of specific communities. Within the context of the scientific community, for instance, they could specify norms of citation and attribution relevant to their discipline, such as archaeology or biology.

The use of a non-binding and flexible Community Norms statement forms part of the Science Commons Protocol by not creating a strict legal obligation for the data.

CCZero

Creative Commons (CC) has also implemented the Science Commons Protocol with their own public domain tool, CCZero, based in part on their earlier work on the Public Domain Dedication (http://creativecommons.org/license/publicdomain-2) currently available on the CC site. CCZero is at the same time an implementation of the Protocol for data and an expanded and clarified version of their public domain dedication. The CCZero tool applies to all types of content, not just data. The following comments are based on the beta CCZero tool (http://labs.creativecommons.org/license/zero) available for comment and discussion at the time of this writing.

CCZero uses two underlying legal tools: one waives rights (http://labs.creativecommons.org/licenses/zero-waive/1.0/us/legalcode) and the other asserts that rights do not exist (http://labs.creativecommons.org/licenses/zero-assert/1.0/us/legalcode). The waiver works for authors and rights holders and the assertion means that someone believes that the work has no copyright in the United States.

Both variations of the CCZero tool are based on US law, and CC anticipates that CCZero will continue to be "ported" to jurisdiction-specific implementations (http://creativecommons.org/international/) via their international affiliates. The Canadian version is available from http://creativecommons.org/international/ca/.

In comparison to CCZero, the Open Data Commons PDDL is:

• based exclusively around databases and data

• international in scope in one single document

• integrates seamlessly with the Community Norms statement

These differences primarily arise from the different focus and infrastructure of the ODC project, though both projects implement the Science Commons Protocol.

The Open Data Commons PDDL interoperates with CCZero via the CCZero assertion tool. Under the current CCZero beta tool, users go through a point and click process and, when prompted, enter in the reason why they believe that the information covered by CCZero has public domain status. At this point, the user can indicate that the PDDL covers the work as the reason for public domain status. This way, ODC can take advantage of the framework being developed for CCZero and the high profile of Creative Commons licensing activities. Users are not confronted with stand-alone licence silos where information covered by one licence cannot be integrated with information under another licence: the PDDL and CCZero fully integrate.

Final Thoughts

The end result of the Science Commons Protocol and the implementation by ODC are solutions for those wishing to further data integration projects and to openly share their data. The PDDL together with the accompanying Community Norms statement will be particularly useful for scientists wishing to share their research data.

But scientists are not the only anticipated users, as government sector data services, and private companies involved in data generation and sharing will all have an interest — as both consumers and producers of data — in having an option that allows for use and re-use of databases without restriction. The goal of the ODC project is to grow with the support of its users to meet the need for accessible legal tools for the creation of a web of open data of all types.

Should you wish to support the ODC's efforts to create data licensing solutions either financially or with your time, please contact
support@opendatacommons.org.

Jordan Hatcher helps people understand intellectual property and Internet law, especially issues surrounding open licensing solutions such as Creative Commons and open data. He has a JD in law from the University of Texas, and a LLM in IP and IT law from the University of Edinburgh, Scotland. He is also the author, together with Dr. Charlotte Waelde, of the Open Data Commons set of legal tools. You can find out more about him at
http://opencontentlawyer.com/.

Recommended Resources

Science Commons Protocol for implementing Open Access Data
http://sciencecommons.org/projects/publishing/open-access-data-protocol/

Science Commons Database Protocol FAQ
http://sciencecommons.org/resources/faq/database-protocol/

"So imagine a mirror
Bigger than the room it was placed in
Imagine my wish for a future
that cannot hold my wish
Imagine the want to hold a rod
that cannot hold the fish
Imagine a rod that cannot hold the fish"
　　　　　Paul Heaton, The Beautiful South

Digital technologies have forever changed the way that knowledge is disseminated and accessed. Yet, the main problem knowledge workers face is invisibility: if people don't know that you know, and people are not aware of what you know, you do not exist.

Governments and institutions are being pushed to foster Open Access (OA) literature as a way to achieve universal reach of research diffusion at inexpensive and immediate levels. Most efforts have been made at the institutional level, dedicating little energy to what the individual can do to contribute. The philosophy and tools around web 2.0 bring clear opportunities for individuals to contribute and to build a broader personal presence on the Internet and a better diffusion for their work, interests or publications.

We propose the concept of the personal research portal (PRP) as a means to create a digital identity for knowledge workers – tied to one's digital public notebook and personal repository – and a virtual network of colleagues working in the same field. Complementary to formal publishing or taking part in offline meetings, the PRP would be a knowledge management system that would enhance reading, storing, and creation at both the private and public levels, and contribute to create an online identity that, in turn, will help to create a network whose currency is knowledge.

Beyond the e-Portfolio

The approach we present is closely related to the concept of e-portfolio (http://en.wikipedia.org/wiki/E-portfolio), but from the researcher's or practitioner's point of view rather than the learner's. As e-portfolios are usually associated with students and teaching, we here propose the term personal research portal (PRP) to avoid confusion.

The goals of the PRP should be: i) to gain more access to international knowledge output from other knowledge workers; ii) to give more international access to research generated by a knowledge worker; iii) higher promotion of institutional knowledge output; iv) improved citation and research impact; v) improved access to subsidiary data; and vi) a strongly facilitated peer review.

To do so, the PRP should be a low cost, highly flexible virtual space in order to:

• host a public repository for interlinked personal production which includes past and present (work in progress) information and documentation

• gather digital resources, news, general information and materials which are accessible from each and every computer

• self-archive and self-publish ongoing research while avoiding waits and delays

• increase one's visibility while enabling networking and knowledge sharing

All in all, the PRP should track the read-think-write routine performed by practitioners and scientists involved in research.

The big difference from publishing is that the PRP should not only keep record of stock knowledge, formal knowledge that lasts or should last, but also flow knowledge, non-structured knowledge that is not intended to remain permanent because it is devoted to foster exchange.

While many PRPs could potentially be built from a mesh of different applications, we here propose a prototype built from these components:

• a static web site with personal and professional information drawing the researcher profile

• a blog for recording news, reflections, and flow knowledge arising from readings, research results and hypotheses

• a blogroll to provide both a live reader for the researcher and a live bibliography of bookmarks for the community

• a wiki to store stock knowledge which evolves over time with the collaboration of third parties

• a bibliographic manager with online access to all or most records

• a personal repository to self-archive published papers as well as self-publish preprints, working papers, presentations, and syllabuses

• social bookmarking tools and file stores for image, sound, and video

• RSS feeds for every dynamic page

Social Software

In the last few years, new and easy to use web tools became available which provide interconnectivity allowing for communication and collaboration, with the only pre-requisite being a personal computer connected to a network.

These technologies represent inexpensive, highly accessible means allowing anyone to share their knowledge with their peers. At least three immediate reflections arise:

• they provide a way for people to easily share, publicize, and diffuse their findings as well as who is behind them

• information published in this manner is easily available from anywhere

• the more everyone engages in a community, the richer it becomes

A possible barrier for using such applications is capacity building in users. Web 2.0 applications and social software are designed for non-technical users, thus only a low level of digital literacy is required. From our point of view, the major requirements to enter the conversation are some degree of e-awareness, a minimum of digital culture, and the ability to change.

To overcome this last barrier, we believe that the network of peers itself, boosted by social software, can help stewarding technology. Wikis, forums, blogs, and other tools provide perfect companions to take the newcomer by the hand in his way into web 2.0. Of course, the conversation is also affected by different cultural backgrounds and different mother tongues, but this is not exclusive from online interaction and, moreover, local communities can form without the mediation of formal literature.

A PRP Prototype

Here we suggest a prototype of a PRP in order to provide a background image of reference. Our philosophy in building is not coding from scratch, but combining existing tools.

We also think that mastering some of these tools will soon become basic literacy skills, much as typewriting, writing an essay, or imparting a live presentation.

Our first consideration would be obtaining one's own domain and hosting. The former, because a domain name is automatically associated with a specific content and its managers. The latter, to retain autonomy of the services, shape, and content on the site.

Static pages and most of the dynamic ones can be built using an open source content management system (CMS) such as Drupal (http://drupal.org/) or Joomla (http://www.joomla.org/). WordPress (http://wordpress.org/) is a blog engine that can also be used as a CMS. Examples of e-portfolio applications include Elgg (http://elgg.org/) and OSPI (http://www.theospi.org/).

With regards to collaboration: if the expected output is content, a wiki is the best option. If the goal is the process and the debate itself, then forums are required. Some of the preceding applications include wikis and message boards. Mediawiki (http://mediawiki.org/) for the wiki, and phpBB (http://phpbb.com) for the message board, are also good choices.

Concerning bibliographic managers, Refbase (http://refbase.sourceforge.net/) and BibCiter (http://bibciter.sourceforge.net/) are web based and provide RSS output. EPrints (http://www.eprints.org/) and Open Journal Systems (http://pkp.sfu.ca/ojs/) provide for self-archiving and self-publishing, respectively.

There are many other applications to share bookmarks, photos and slideshows, to publish podcasts or vodcasts.

However, most are online services which are provided and hosted by third parties. Their use should be based on the availability to import and export one's data and should be properly linked on the PRP. RSS output, the glue, is a must.

When connectivity is not available and working locally should be made possible, XAMPP (http://sourceforge.net/projects/xampp/) provides the ability to (re)install all the social software applications to the local hard drive or a USB pen drive. Indeed, it can work as a backup for the PRP as well as make it portable across different operating systems.

Digital Identity

In an age of information overload, one of the main problems that knowledge workers face is invisibility. This invisibility causes, at least, two major consequences:

• minimum awareness and recognition of one's findings, fields of work, interests and even existence

• difficult access to mainstream circuits in one's field

It is thus important that researchers gain visibility so that they and their work become known in academic and practitioner circles at the international level.

Setting up a PRP should hence be understood, at a primary level, as the creation of a personal home page. Notwithstanding, this digital identity, or the researcher's presence on the Internet, is juxtaposed to the identity shown by authorship in paper journals and conference speeches, with each identity complementing the other. While the latter identity is strongly tied to a handful of concepts exposed in a determinate paper, the digital identity should give further information on the following aspects:

• who and where am I?

• what do I do?

• what interests me?

• what have I done?

If mainstream systems such as congresses, journals, and seminars act as diffusion hubs for offline identities, search engines, portals, blogs, institutional pages, and signature files in e-mails act as diffusion hubs for online identities.

Nevertheless, there are, in our opinion, two main differences among both channels: the higher potential reach of online media and the always updated information provided by PRPs, especially those provided with an RSS feed and correctly meta-tagged data.

Summing up, the main component of a PRP should be evolving, up-to-date information of one's work. Search engines are web 2.0-friendly and highly score live pages with rich and focused content. Descriptions about one's research and interests, side-by-side to documents and other materials and links to and from other people with similar interests enhances the possibility of being found under determinate keywords. This information should be created through static pages by means of simple HTML documents or, better, using a CMS or CMS-like features from other applications such as blogs. The blogroll should play, among others, a great part in the linking role.

Reinforcing Digital Identity

In the process of gathering information to increase one's knowledge and prepare research, it is usual to take notes, highlights of what has been read, reflections that arise after the reading, or just a notation of the fact that something has been read.

Social software empowers researchers as their notes can be published automatically and, even more, link the people and documents that generated these notes. Thus, the digital notebook makes the read, write, analyze, reflect, and learn process fully public. Another immediate consequence is that a live digital store is created daily, a store that is categorized, searchable and fully accessible, with absolute immediacy and no filtering other than your own criterion.

Joining the blog as a collector of flow knowledge, a wiki allows all sorts of content interlinking, tagging and categorizing with the aim of increasing the information available as a whole and enabling collaboration.

A last tool worth looking at is a bibliographic manager. The evident use of a bibliographic manager is keeping all one's references properly sorted and, in some cases, providing tools to ease the task of citing while writing. Some are web applications installed on a web server which allows not only managing but publishing one's references and bibliographies. This feature contributes to both building one's digital identity, by wrapping all the PRP with names and references belonging to the same area of knowledge that reinforce the identity's framework, as well as shared content to the PRP, which was one of the main goals of the PRP. It is also more attractive to Internet search engines, again reinforcing the achievement of the visibility goal.

Network Building

After identity, meeting other colleagues, exchanging impressions, and working together is what social software is all about.

We want to stress the point that, more than search engines, RSS feeds enable knowledge sharing in real time.

RSS feeds allow subscription, that can be selective through tags, syndication and aggregation to new knowledge created around the world.

Fostering community building will be enhanced by citations and their corresponding links, pingbacks and trackbacks. This interaction can be reinforced by comments on others' PRPs or the creation of friend of a friend (FOAF, http://www.foaf-project.org/) files and blogrolls. These last two shape a virtual research network around the PRP and its creator. The extension of this behavior among other researchers helps invisible knowledge workers become present in the relevant, virtual forums.

The PRP reduces contact time as one is findable, and can enrich this contact time because all the information is already there for anyone to read, thus enabling peer review. A higher exposure allows for more highly informed dialogues to take place, paves the path to future collaborations, and shifts a cultural change towards openness.

Of course, no conversation takes place by only speaking, so a feedreader will also become a perfect companion to one's blog.

Self-archiving, Self-publishing

We have talked so far about virtual identity, the digital notebook, the collection of content and explicit knowledge, and the creation of networks. We should not forget that sometimes we create knowledge that should be published, not as notes, but as a finished work.

The PRP provides for self-archiving one's preprints and published works in a personal repository. This does not solve the problem of access to journal publishing itself, but it does solve access to published works.

The author can self-publish and obtain an ISSN or ISBN for each published work within the PRP. Such publications contribute to increase the visibility of the author, shape his digital identity, enrich the content of the site, and make it more appealing to users and search engines.

Conclusions

While increasing, there is still underuse of wikis, social bookmarking, social networking, file sharing, RSS feeds, discussion forums and blogs within academic circles. Researchers and practitioners, faculty and non-scholars, experts and learners, managers and engineers have the ability to provide plenty of knowledge in their lives and works. If shared, this knowledge will be a part of a network worth keeping. If not, these circles will be disconnected and starve. The PRP could help these knowledge workers both as a personal knowledge manager and as a rich knowledge network weaver. Costs are few and benefits are many.

This article is based on the author's paper: The Personal Research Portal: Web 2.0 Driven Individual Commitment with Open Access for Development (http://www.km4dev.org/journal/index.php/km4dj/article/view/92).

Ismael Peña-López is Public Policies for Development and ICT4D lecturer at the Open University of Catalonia, Spain. His main research interests are the digital divide, e-readiness, ICT4D and digital inclusion. As a teacher in virtual learning environments, he is also interested in Open Access and its relationship with e-learning and development. He is the lead (and only) developer of Bibciter, a GPL Bibliography Manager. His Personal Research Portal can be accessed at ICTlogy (http://ictlogy.net/).

The goal of the Talent First Network Proof of Principle (TFN-POP) is to establish an ecosystem anchored around the commercialization of open source technology developed at academic institutions in Ontario.

The priority areas are the commercialization of open source in:

• Mapping and geospatial applications

• Simulation, modeling, games, and animation

• Conferencing

• Publishing and archiving

• Open educational resources

• Social innovation

• Business intelligence

• Ecosystem management

• Requirements management

Expected Results

The TFN-POP is expected to:

• Establish a healthy ecosystem anchored around the commercialization of open source assets

• Maximize the benefits of the investment in the Talent First Network by the Ministry of Research and Innovation

• Accelerate the growth of businesses in Ontario that use open source assets to compete

Eligibility to Receive Funds

Individuals eligible to receive funds are faculty, staff, and students of universities and colleges in Ontario.

Budget and Size of Grants

A total of $300,000 is available. Applicants' requests should not exceed $30,000.

The TFN-POP may provide up to 50 percent of total project costs.

Criteria

Proposals will be judged against the following five criteria:

• Strength and novelty of open source technology proposed

• Extent of market advantage due to open source

• Project deliverables, likelihood that the proposed activities will lead to deliverable completion on time, and effectiveness of the plan to manage the project

• Track record and potential of applicants

• Extent of support from private sector

Application

The electronic version of the application received by email at the following address: TFNCompetition@sce.carleton.ca will be accepted as the official application. The email must contain three documents: a letter of support, project's vitals, and a project proposal.

Letter of support: (maximum 2 pages) a letter, signed by the person responsible for the Technology Transfer Office or Applied Research Office of the academic institution that proposes to host the project and the faculty member or student who will lead the project, must be included. This letter should describe the nature of the support for the project from the academic institutions, companies and other external organizations.

Project's vitals: (maximum 1 page) The project's vitals must include:

• Person responsible for applied research or technology transfer at the college submitting the proposal: name, mailing address, telephone number, and email address

• Project leader: name, mailing address, telephone number, and email address

• Team members: names, mailing addresses, telephone numbers, and email addresses

• Budget: Total budget, with TFN's contribution and that of other organizations

• TFN investment: TFN contribution broken down by payments to students, payments to faculty, and payments to project awareness activities

Project proposal: (maximum 5 pages) Project proposal must include the following:

• Benefits: (maximum 1/2 page) Description of the benefits of the proposed project, and an overview of the context within which the project is positioned

• Advantage: (1/2 page) Market advantage provided by open source assets used in the project

• Information on applicants: (maximum 1.5 pages) Background information to help assess the track record and potential of the people who are key to the project and the college

• Project plan: (maximum 2.5 pages) Description of the deliverables (what will be delivered and when); key project activities; nature of the involvement from companies, and other external organizations; and plan to manage the project

Evaluation & Deadline

Proposals will undergo review by the Expert Panel established by the TFN-POP. The Chair of the Panel may contact the applicants if required. A final decision will be communicated to the applicants within 30 days after the email with the official application is received.

There is no deadline. Applications will be evaluated on a first-come basis until the $300,000 available is committed.

Contacts

Luc Lalande: Luc_Lalande@carleton.ca

Rowland Few: rfew@sce.carleton.ca

About the Talent First Network

The Talent First Network (TFN) is an Ontario-wide, industry driven initiative launched in July 2006 with the support of the Ministry of Research and Innovation and Carleton University. The objective is to transfer to Ontario companies and Open source communities: (i) Open source technology, (ii) knowledge about competing in Open source environments and (iii) talented university and college students with the skills in the commercialization of Open source assets.

Innovativeness of Software Solutions: Comparing Free/Open Source and Proprietary Products

From the Abstract:

The issue of innovation processes taking place in the software sector is currently widely debated. Challenging questions arise about what products/services have to be considered innovative, and whether a specific artefact is innovative or not. In this framework, the widespread success of the Free/Open Source Software (FOSS) put forward new research issues, dealing with whether and how programs developed according to the new production paradigm turn out to be more innovative than traditional ones. In this framework, this paper aims at contributing to the literature by addressing three main research questions: (i) are software solutions produced by Small and Medium Enterprises (SMEs) innovative? (ii) What kinds of innovations are implemented? And, finally, (iii) are programs based on FOSS more innovative than proprietary ones?

http://opensource.mit.edu/papers/lorenzi_rossi_MIT_20071220.pdf

Involving Software Engineering Students in Open Source Software Projects

From the Abstract:

Anecdotal and research evidences show that the Free and Open Source Software (F/OSS) development model has produced a paradigm shift in the way we develop, support, and distribute software. This shift is not only redefining the software industry but also the way we teach and learn in our software engineering (SE) courses. But for many universities F/OSS is seen as an optional low cost technology to support the IT infrastructure and administrational duties. Few see F/OSS as an opportunity for students to learn the SE concepts and skills we teach. Furthermore, it is still an open question as to whether the F/OSS methodology can be effectively used to teach SE courses within the formally structured curriculum in most universities. This paper discusses F/OSS projects as bazaars of learning that offer a meaningful learning context.

http://opensource.mit.edu/papers/bazaars_of_learning.pdf

March 12-14

PHP Quebec

Montreal, QC

PHP Quebec is pleased to present the sixth edition of the PHP Quebec Conference. PHP experts will be presenting real life solutions to developers and project managers. With events such has the PHPLab, the Job Fair and the Cocktail, the Conference will be a great opportunity to meet with local experts, core PHP developers and sponsors.

http://conf.phpquebec.com/en

April 7-9

IT360

Toronto, ON

"Cultivating Innovation in Technology" takes off at the interactive IT360° experience. IT360° is the only industry event that fully integrates key industry participants from diverse sectors. IT360° is a "teaching conference" where you will learn about current applications and solutions, innovations, tools & technologies, what works and what does not and what is most relevant to your organization. Content areas include open source, security, SOA, IT green, and storage.

http://www.it360.ca/

April 10

The Entrepreneur: Heretic or Hero of Innovation?

Ottawa, ON

This presentation will address a number of areas and lessons including: negotiating some early intellectual property from NRC, inventing new technology and the importance of patents for a start-up, pros and cons of raising funding from Angel Investors, VCs, government programs, and building out the team. Pre-registration is mandatory for this Colloquium Series event.

http://iit-iti.nrc-cnrc.gc.ca/colloq/0708/08-04-10_e.html

31

The goal of the Open Source Business Resource is to provide quality and insightful content regarding the issues relevant to the development and commercialization of open source assets. We believe the best way to achieve this goal is through the contributions and feedback from experts within the business and open source communities.

OSBR readers are looking for practical ideas they can apply within their own organizations. They also appreciate a thorough exploration of the issues and emerging trends surrounding the business of open source. If you are considering contributing an article, start by asking yourself:

1. Does my research or experience provide any new insights or perspectives?

2. Do I often find myself having to explain this topic when I meet people as they are unaware of its relevance?

3. Do I believe that I could have saved myself time, money, and frustration if someone had explained to me the issues surrounding this topic?

4. Am I constantly correcting misconceptions regarding this topic?

5. Am I considered to be an expert in this field? For example, do I present my research or experience at conferences?

If your answer is "yes" to any of these questions, your topic is probably of interest to OSBR readers.

When writing your article, keep the following points in mind:

1. Thoroughly examine the topic; don't leave the reader wishing for more.

2. Know your central theme and stick to it.

3. Demonstrate your depth of understanding for the topic, and that you have considered its benefits, possible outcomes, and applicability.

4. Write in third-person formal style.

These guidelines should assist in the process of translating your expertise into a focused article which adds to the knowledgable resources available through the OSBR.

Upcoming Editorial Themes	
March 2008	Procurement
April 2008	Communications
May 2008	Enterprise Readiness
June 2008	Security
July 2008	Accessibility

Formatting Guidelines:

All contributions are to be submitted in .txt or .rtf format and match the following length guidelines. Formatting should be limited to bolded and italicized text. Formatting is optional and may be edited to match the rest of the publication. Include your email address and daytime phone number should the editor need to contact you regarding your submission. Indicate if your submission has been previously published elsewhere.

Articles: Do not submit articles shorter than 1500 words or longer than 3000 words. If this is your first article, include a 50-75 word biography introducing yourself. Articles should begin with a thought-provoking quotation that matches the spirit of the article. Research the source of your quotation in order to provide proper attribution.

Interviews: Interviews tend to be between 1-2 pages long or 500-1000 words. Include a 50-75 word biography for both the interviewer and each of the interviewee(s).

Newsbytes: Newsbytes should be short and pithy--providing enough information to gain the reader's interest as well as a reference to additional information such as a press release or website. 100-300 words is usually sufficient.

Events: Events should include the date, location, a short description, and the URL for further information. Due to the monthly publication schedule, events should be sent at least 6-8 weeks in advance.

Questions and Feedback: These can range anywhere between a one sentence question up to a 500 word letter to the editor style of feedback. Include a sentence or two introducing yourself.

Copyright:

You retain copyright to your work and grant the Talent First Network permission to publish your submission under a Creative Commons license. The Talent First Network owns the copyright to the collection of works comprising each edition of the OSBR. All content on the OSBR and Talent First Network websites is under the Creative Commons attribution (http://creativecommons.org/licenses/by/3.0/) license which allows for commercial and non-commercial redistribution as well as modifications of the work as long as the copyright holder is attributed.

The Talent First Network program is funded in part by the Government of Ontario.

The Technology Innovation Management (TIM) program is a master's program for experienced engineers. It is offered by Carleton University's Department of Systems and Computer Engineering. The TIM program offers both a thesis based degree (M.A.Sc.) and a project based degree (M.Eng.). The M.Eng is offered real-time worldwide. To apply, please go to: http://www.carleton.ca/tim/sub/apply.html.